AD-A119 609 EDUCATIONAL TESTING SERVICE PRINCETON NJ F/G 12/1
SAMPLING VARIANCES AND COVARIANCES OF PARAMETER ESTIMATES IN IT--ETC(U)
AUG 82 F M LORD, M.S WINGERSKY N00014-80-C-0402
UNCLASSIFIED ETS-RR-82-33-ONR NL

1 OF 1
ADA
119609

END
DATE
FILMED
11-82
DTIC

AD A119609

(12)

RR-82-33-ONR

SAMPLING VARIANCES AND COVARIANCES OF PARAMETER ESTIMATES IN ITEM RESPONSE THEORY

Frederic M. Lord

and

Marilyn S. Wingersky

This research was sponsored in part by the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research, under
Contract No. N00014-80-C-0402

Contract Authority Identification Number
NR No. 150-453

Frederic M. Lord, Principal Investigator

Educational Testing Service

Princeton, New Jersey

August 1982

SAMPLING VARIANCES AND COVARIANCES OF PARAMETER ESTIMATES IN ITEM RESPONSE THEORY

Frederic M. Lord

and

Marilyn S. Wingersky

This research was sponsored in part by the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research, under
Contract No. N00014-80-C-0402

Contract Authority Identification Number
NR No. 150-453

Frederic M. Lord, Principal Investigator

Educational Testing Service

Princeton, New Jersey

August 1982

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A119609	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Sampling Variances and Covariances of Parameter Estimates in Item Response Theory		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER RR-82-33-ONR
7. AUTHOR(s) Frederic M. Lord and Marilyn S. Wingersky		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0402
9. PERFORMING ORGANIZATION NAME AND ADDRESS Educational Testing Service Princeton, NJ 08541		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 150-453
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE August 1982
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Sampling Covariances
Standard Errors
Item Response Theory
Maximum Likelihood Estimators
Item Parameters
Matrix Inversion

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper develops a possible method for computing the asymptotic sampling variance-covariance matrix of joint maximum likelihood estimates in item response theory when both item parameters and abilities are unknown. For a set of artificial data, results are compared with empirical values; also with the variance-covariance matrices found by the usual formulas for the case where the abilities are known, or where the item parameters are known. The results are consistent with the conjecture that the new method is asymptotically correct except for errors due to grouping.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Sampling Variances and Covariances

of Parameter Estimates in Item Response Theory

Abstract

This paper develops a possible method for computing the asymptotic sampling variance-covariance matrix of joint maximum likelihood estimates in item response theory when both item parameters and abilities are unknown. For a set of artificial data, results are compared with empirical values; also with the variance-covariance matrices found by the usual formulas for the case where the abilities are known, or where the item parameters are known. The results are consistent with the conjecture that the new method is asymptotically correct except for errors due to grouping.

Sampling Variances and Covariances

of Parameter Estimates in Item Response Theory*

In item response theory (IRT), the observations come in the form of an n -by- N matrix, with one row for each item and one column for each examinee. The joint frequency distribution of the observations depends on a vector of N 'ability' parameters--one for each person--and on a matrix of item parameters. Here, we will consider only the three-parameter logistic model for dichotomously scored items, so there will be three item parameters (a , b , and c) for each of n items. A method will be developed for computing the asymptotic sampling variance-covariance matrix when both abilities and item parameters are unknown. Until this is done we do not know the standard errors of the parameter estimates, which handicaps development of a goodness-of fit test and other statistics required in applications of IRT.

If the item (ability) parameters are known, the estimated ability (item) parameters have independent sampling distributions. It can be shown (see Bradley & Gart, 1962) that the maximum likelihood estimates of the ability (item) parameters are consistent. Hence the asymptotic sampling variance for an estimated ability parameter is given by the usual formula

$$\mathrm{Var}(\hat{\theta}_r \mid a,b,c) = [\mathcal{E}(\partial \ell / \partial \theta_r)^2]^{-1} \quad , \qquad (1.1)$$

where $\hat{\theta}_r$ is the estimated ability parameter, ℓ is the log of the likelihood, and a , b , and c are the known vectors of item parameters.

---

*This work was supported in part by contract N00014-80-C-0402, project designation NR 150-453 between the Office of Naval Research and Educational Testing Service. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Similarly the asymptotic sampling variance-covariance matrix of the estimated item parameters for an item is given by

$$\|\mathrm{Cov}(\hat{\tau}_v, \hat{\tau}_w \mid \theta)\| = \left\| \mathcal{E}\left(\frac{\partial \ell}{\partial \tau_v} \frac{\partial \ell}{\partial \tau_w} \right) \right\|^{-1} \qquad (v,w = 1,2,3) \qquad (1b)$$

where $\{\tau_v\}$ is a vector consisting of the estimated a , b , and c for a single item and θ is the known vector of abilities. The right-hand side is the inverse of a 3-by-3 matrix.

When neither item nor ability parameters are known, all parameters are often estimated simultaneously by maximum likelihood. In the (Rasch) case where there is only one parameter per item, Haberman (1977) has shown that all parameter estimates will converge to their true values (will be consistent) when the number of examinees and the number of test items become large simultaneously. Empirical results suggest that consistency probably also holds when all parameters are estimated simultaneously under the three-parameter model. If so, it is reasonable that the asymptotic sampling variance-covariance matrix of all estimated parameters will be given by the usual formula

$$\|\mathrm{Cov}(\hat{\tau}_p, \hat{\tau}_q)\| = \left\| \mathcal{E}\left(\frac{\partial \ell}{\partial \tau_p} \frac{\partial \ell}{\partial \tau_q} \right) \right\|^{-1} \qquad (p,q = 1,2,\ldots,M) \qquad (2)$$

where $M = 3n + N - 2$ and $\tau \equiv \{\tau_p\} \equiv \{a_1, b_1, c_1, a_2, b_2, c_2, \ldots, a_n, b_n, c_n; \theta_1, \theta_2, \ldots, \theta_{N-2}\}$.

Since standard errors are urgently needed in practical work where all parameters are estimated simultaneously by maximum likelihood, this report compares numerical values provided by (2) with values provided by (1) and with empirically observed sampling fluctuations. The comparisons to be presented suggest that (2) provides useful values for the desired standard errors.

There are several special problems that arise in the evaluation and practical utilization of (2), problems that do not arise in the situation where (1) is appropriate:

1. Until an origin and scale are specified, the parameters are not identifiable.
2. The mathematical formulation is complicated by the choice of origin and scale.
3. The usual choice of origin and scale when estimating IRT parameters is inconvenient for mathematical purposes.
4. The numerical values of the sampling variances are very much affected by the choice of origin and scale.
5. Equation (2) requires the inversion of a matrix of order $N + 3n - 2$ where N may be several thousand.

These problems will be considered in subsequent sections.

1. Parameterization

The appropriate likelihood function is (Lord, 1980)

$$L(a,b,c;\theta|U) = \prod_{i=1}^{n} \prod_{a=1}^{N} P_{ia}^{u_{ia}} Q_{ia}^{1-u_{ia}} \qquad (3)$$

where θ is the vector of the N ability parameters; a , b , and c are each a vector of n item parameters, $U \equiv \|u_{ia}\|$ is the matrix of item responses u_{ia} (= 0 or 1); finally $Q_{ia} \equiv 1 - P_{ia}$ and P_{ia} is the item response function, the probability of a correct answer by examinee a to item i . Each given P_{ia} is a function of θ_a and of a_i , b_i , and c_i , but not of any other parameters. In numerical work here, P_{ia} will be taken to be the three-parameter logistic function

$$P_{ia} \equiv c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta_a - b_i)]} \quad . \tag{4}$$

For mathematical purposes, however, it is only necessary to state that P_{ia} is an increasing function of θ_a .

If we add some constant to all θ_a and subtract the same constant from all b_i , all P_{ia} will be unchanged. This means that the origin used for measuring ability is entirely arbitrary. If we multiply each θ_a and each b_i by some constant and divide each a_i by the same constant, again all P_{ia} will be unchanged. This means that the unit used to measure ability is entirely arbitrary. Since we can change the origin and unit of the θ_a without changing (3), it follows that θ , a , b , and c are not identifiable and cannot be estimated from (3) without further specification.

To conform to a commonly used procedure, we could choose the origin and scale so that for some specified group of examinees the mean of the θ_a is zero and the variance is one. This is not convenient mathematically, however. Instead, two other methods of

specifying the origin and scale will be used, even though this will complicate matters later on when the results are applied in practice. In the first method, without loss of generality, arbitrary numerical values will be assigned to θ_{N-1} and to θ_N.

The $M \equiv N + 3n - 2$ likelihood equations are

$$0 = \sum_{i=1}^{n} \sum_{a=1}^{N} (u_{ia} - P_{ia}) \frac{P_p^{ia}}{P_{ia}Q_{ia}} \qquad (p = 1,2,\ldots,M) \tag{5}$$

where $P_p^{ia} \equiv \partial P_{ia}/\partial \tau_p$.

2. Fisher Information Matrix

The Fisher information matrix on the right of (2) now has as a typical element

$$I_{pq} \equiv \mathcal{E}\left(\frac{\partial \ell}{\partial \tau_p} \frac{\partial \ell}{\partial \tau_q} \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{a=1}^{N} \sum_{b=1}^{N} \frac{P_p^{ia} P_q^{jb}}{P_{ia}Q_{ia}P_{jb}Q_{jb}} \operatorname{Cov}(u_{ia}, u_{jb})$$

$$(p,q = 1,2,\ldots,M) \quad .$$

Because of local independence and random sampling of examinees,

$$\operatorname{Cov}(u_{ia}, u_{jb}) = \delta_{ij}\delta_{ab}P_{ia}Q_{ia}$$

where $\delta_{st} = 1$ if $s = t$, $\delta_{st} = 0$ otherwise. Thus the typical element is

$$I_{pq} \equiv \sum_{i=1}^{n} \sum_{a=1}^{N} \frac{P_p^{ia} P_q^{ia}}{P_{ia}Q_{ia}} \qquad (p,q = 1,2,\ldots,M) \quad . \tag{6}$$

Note that P_p^{ia} is zero unless either p and a refer to the same person, or p and i refer to the same item. Thus

$$
\|I_{pq}\| = \left[\begin{array}{cccc|cccc}
S_1 & 0 & \cdots & 0 & \underset{\sim}{f}_{11} & \underset{\sim}{f}_{12} & \cdots & \underset{\sim}{f}_{1N'} \\
0 & S_2 & \cdots & 0 & \underset{\sim}{f}_{21} & \underset{\sim}{f}_{22} & \cdots & \underset{\sim}{f}_{2N'} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & S_n & \underset{\sim}{f}_{n1} & \underset{\sim}{f}_{n2} & \cdots & \underset{\sim}{f}_{nN'} \\
\hline
\underset{\sim}{f}'_{11} & \underset{\sim}{f}'_{21} & \cdots & \underset{\sim}{f}'_{n1} & t_1 & 0 & \cdots & 0 \\
\underset{\sim}{f}'_{12} & \underset{\sim}{f}'_{22} & \cdots & \underset{\sim}{f}'_{n2} & 0 & t_2 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
\underset{\sim}{f}'_{N'1} & \underset{\sim}{f}'_{N'2} & \cdots & \underset{\sim}{f}'_{N'n} & 0 & 0 & \cdots & t_{N'}
\end{array}\right] \tag{7}
$$

where $N' \equiv N - 2$, S_i is the 3-by-3 Fisher information matrix for a_i, b_i, and c_i, t_a is the Fisher information for examinee a, and $\underset{\sim}{f}_{ia}$ is the 3-by-1 joint Fisher information vector for item i and examinee a:

$$
\underset{\sim}{f}_{ia} \equiv \frac{\partial P_{ia}/\partial \theta_a}{P_{ia} Q_{ia}} \begin{bmatrix} \partial P_{ia}/\partial a_i \\ \partial P_{ia}/\partial b_i \\ \partial P_{ia}/\partial c_i \end{bmatrix} .
$$

3. Matrix Inversion

The following general formula for inverting a partitioned matrix may be applied to (7)

$$\left[\begin{array}{c|c} S & F \\ \hline F' & T \end{array}\right]^{-1} \equiv \left[\begin{array}{c|c} S^{-1} + S^{-1}FZ^{-1}F'S^{-1} & -S^{-1}FZ^{-1} \\ \hline -Z^{-1}F'S^{-1} & Z^{-1} \end{array}\right] \tag{8}$$

where

$$Z \equiv T - F'S^{-1}F \quad . \tag{9}$$

The matrix S is easily inverted since it is a diagonal supermatrix:

$$S^{-1} = {}_{o}\Big\| \; S_i^{-1} \; \Big\|^{o} \quad .$$

The notation on the right denotes a diagonal matrix with diagonal elements S_i^{-1} . These last are easily computed since each S_i is only 3 by 3.

All the matrix operations indicated on the right side of (8) can be carried out on the computer without difficulty, with one exception: the inversion of Z , which is N' by N' . The approximation used here to invert Z relies on grouping the θ_a into 16 class intervals of width 0.5, covering the range $-5 \le \theta_a \le 3$. Each θ_a in a given class interval is replaced by the midpoint of the interval.

Now T will be a diagonal supermatrix $T \equiv \|T_g\|$, where $T_g \equiv t_g I$ is a scalar matrix with dimensions N_g by N_g, and N_g is the number of people in class interval g. Also, F will be a row vector of 16 matrices, the columns of any one matrix being all identical:

$$F \equiv \{f_1 1_1', f_2 1_2', \ldots, f_{16} 1_{16}'\} \quad , \tag{10}$$

where $f_g \equiv \{f_{ia}\}$ for any examinee a in class interval g and 1_g is a unit vector whose length is N_g.

The product $F'S^{-1}F$ can now be written as a 16-by-16 supermatrix:

$$F'S^{-1}F = \|1_g f_g' S^{-1} f_h 1_h'\| \quad .$$

Denote the scalar $f_g' S^{-1} f_h$ by w_{gh}. We now have

$$Z \equiv T - \|M_{gh}\| \quad , \tag{11}$$

$$M_{gh} \equiv w_{gh} 1_g 1_h' \quad . \tag{12}$$

For computation purposes, Z still has N' rows and columns, not just 16. For the usual sample size, it is still not feasible to invert Z with a standard inversion program.

Consider the problem of inverting Z_{11}, the N_1-by-N_1 upper left corner of Z. By (11), (12), and a standard formula,

$$Z_{11}^{-1} \equiv [T_1 - w_{11}\mathbf{1}_1\mathbf{1}_1']^{-1} \equiv T_1^{-1} + \frac{w_{11}T_1^{-1}\mathbf{1}_1\mathbf{1}_1'T_1^{-1}}{1 - w_{11}\mathbf{1}_1'T_1^{-1}\mathbf{1}_1} \quad . \tag{13}$$

Since $T_1 \equiv t_1 I$, where t_1 is scalar, this becomes

$$Z_{11}^{-1} = \frac{I}{t_1} + \frac{w_{11}\mathbf{1}_1\mathbf{1}_1'}{t_1^2 - t_1 w_{11} N_1} \quad .$$

Next, the upper left 2-by-2 supermatrix in Z can be inverted as in (8), using the standard formula for the inversion of a partitioned matrix:

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}^{-1} = \left[\begin{array}{c|c} Z_{11}^{-1} + Z_{11}^{-1}Z_{12}H^{-1}Z_{21}Z_{11}^{-1} & -Z_{11}^{-1}Z_{12}H^{-1} \\ \hline -H^{-1}Z_{21}Z_{11}^{-1} & H^{-1} \end{array}\right] \tag{14}$$

where $H = Z_{22} - Z_{21}Z_{11}^{-1}Z_{12}$. It can be seen that H has the same general form as Z_{11} and can thus be inverted as in (13); so (14) can readily be calculated.

Next, substitute (14) for Z_{11}^{-1} in the foregoing procedure, and repeat this procedure, in such a way as to invert the upper

left 3-by-3 supermatrix in Z . A total of fifteen repetitions enable us to invert the 16-by-16 supermatrix Z . Equation (8) is now used for one final inversion, the result being the desired variance-covariance matrix of all $N + 3n - 2$ parameters.

The 16-by-16 variance-covariance supermatrix for the $\hat{\theta}_a$ consists of 256 blocks. The elements are all the same within a block except for diagonal blocks, each of which has a variance (instead of a covariance) repeated along its diagonal. Any two examinees in the same class interval will have identical Var $\hat{\theta}$ and identical sampling covariances with any other given parameter estimate.

4. Reparameterization

In Section 1, in order to have identifiable parameters, an origin and scale was chosen so that θ_{N-1} and θ_N had arbitrary preassigned values. Any other choice of origin and scale would result in a linear transformation of parameters. The likelihood function would remain unchanged for every pattern of item responses.

The choice of unit (but not the choice of origin) has one completely obvious effect on the sampling errors of parameter estimates. If the unit is changed, the standard errors for the b 's and 's will be multiplied by the ratio of the new scale unit to the old scale unit. The standard errors for the a 's will be divided by this ratio. A second important effect is easily overlooked: the standard error

of the maximum likelihood estimator depends not only on the choice of scale, but also on how the (origin and) scale is specified.

Suppose that the true numerical values of all θ_a (a = 1,...,N) are specified on some arbitrary scale. Suppose next that our test is too difficult for examinee N . This means that the likelihood function is rather insensitive to variations in θ_N . If we could repeat our testing with several parallel test forms, we would find a wide range of estimates of θ_N . In such a situation, the difference between true θ_{N-1} and θ_N clearly cannot be estimated well from the examinee responses. If we define the scale by treating θ_N and θ_{N-1} as known, our estimates of every θ_a may fluctuate grossly, simply because the scale unit $\theta_N - \theta_{N-1}$ is not well determined by the data.

Suppose next that we relabel all examinees so that examinees N - 1 and N are not the same examinees as before. The ability scale has not been changed from the preceding paragraph; it is the procedure for defining the scale that has been changed. The true θ for each examinee is still the same as before. Suppose the new examinees N - 1 and N are both at ability levels where our test measures accurately. If, further, the true θ_{N-1} and θ_N are substantially different from each other, the difficulty of the previous paragraph disappears: Throughout the ability range where the test is designed to measure accurately, the standard errors of all θ_a may be reasonably small.

For example, suppose on some scale $\theta_1 = -3$, $\theta_2 = -2$, $\theta_3 = -1$, $\theta_4 = 0$, $\theta_5 = 1$, $\theta_6 = 2$, $\theta_7 = 3$. We can specify this same scale in terms of any two of these θ 's. The standard errors that we obtain will depend in an overwhelming way not just on the ability scale, but on how we specify it. We cannot rectify the standard errors by some simple procedure, such as multiplying each by a constant.

For this reason, our procedure for specifying the ability scale should depend only on parameters or functions of parameters that are accurately determined by the data. A robust mean of the θ_a might seem attractive; however, any function of the θ_a is counterindicated by the fact that sometimes $\theta_a = \pm\infty$.

The procedure used here is to choose a set of m discriminating, moderately easy items and a set of r discriminating, moderately hard items. We will hereafter define the origin and unit for our new parameters, to be denoted by capital letters, so that the mean of the (true) B -parameters for the easy items is zero, and the mean for the hard items is one.

Our new parameters are related to our old parameters (from Section 2 or from Section 5) by linear transformations:

$$A_i \equiv ka_i \ , \quad B_i \equiv K + b_i/k \ , \quad C_i \equiv c_i \ , \quad \Theta_a \equiv K + \theta_a/k \ , \tag{15}$$

$$(a = 1,2,\ldots,N \ ; \quad i = 1,2,\ldots,n) \ ,$$

where k and K are transformation constants to be determined. Since

$$\bar{B}_0 \equiv \frac{1}{m} \sum^{m} B_i = 0 \ , \quad \bar{B}_1 \equiv \frac{1}{r} \sum^{r} B_i = 1 \quad , \tag{16}$$

the values of k and K are found by substituting (15) into (16) and solving for k and K :

$$k = \bar{b}_1 - \bar{b}_0 \ , \quad K = -\frac{\bar{b}_0}{k} \quad , \tag{17}$$

where $\bar{b}_0$ and $\bar{b}_1$ are means for m and r items, respectively.

To find the variance-covariance matrix for estimates of the upper-case parameters, rewrite (15) as

$$\begin{aligned} &\Theta_a = (\theta_a - \bar{b}_0)/k \ , \quad A_i = ka_i \ , \quad B_i = (b_i - \bar{b}_0)/k \quad , \\ &C_i = c_i \quad . \end{aligned} \tag{18}$$

Because of the special properties of maximum likelihood estimators, equations (18) still hold when estimators are substituted for parameters. Thus the sampling variances and covariances for estimates of the new parameters can be computed from the sampling variances and covariances already obtained at the end of Section 3. Formulas for doing this can be written down from (18) by using the 'delta' method (Kendall & Stuart, 1969, Chapter 10). For example,

$$\left.\begin{aligned}
\mathrm{Cov}(\hat{A}_i, \hat{\theta}_a) &= \mathrm{Cov}(\hat{a}_i, \hat{\theta}_a) - \mathrm{Cov}(\hat{a}_i, \hat{\bar{b}}_0) - \frac{\hat{\theta}_a - \bar{b}_0}{k} \mathrm{Cov}(\hat{a}_i, \hat{k}) \\
&\quad + \frac{a_i}{k} \mathrm{Cov}(\hat{\theta}_a, \hat{k}) - \frac{a_i}{k} \mathrm{Cov}(\hat{\bar{b}}_0, \hat{k}) - \frac{a_i(\theta_a - \bar{b}_0)}{k^2} \mathrm{Var}\ \hat{k} \quad , \\
\mathrm{Cov}(\hat{\bar{b}}_0, \hat{k}) &= \mathrm{Cov}(\hat{\bar{b}}_1, \hat{\bar{b}}_0) - \mathrm{Var}\ \hat{\bar{b}}_0 \quad , \\
\mathrm{Cov}(\hat{\bar{b}}_1, \hat{\bar{b}}_0) &= \frac{1}{mr} \sum^{m} \sum^{r} \mathrm{Cov}(\hat{b}_i, \hat{b}_j) \quad .
\end{aligned}\right\} \quad (19)$$

5. Parameter Estimation

The maximum likelihood estimators (MLE) satisfy the likelihood equations (5). *In* (5), *there is one equation for each parameter* omitting θ_{N-1} and θ_N . If all $N + 3n \equiv M + 2$ MLE are linearly transformed, as for example in (15), the transformed parameters will still satisfy the likelihood equations.

Since the origin and scale for the new parameters is chosen to satisfy (16), then the appropriate k and K are obtained from (17) after replacing $\bar{b}_0$ and $\bar{b}_1$ by their MLE. The likelihood function (3) is unaffected by these linear transformations.

The computer program LOGIST identifies the parameters by still another choice of origin and scale:

1. a certain truncated mean of the $\hat{\theta}_a$ ($a = 1,2,\ldots,N$) is set equal to zero,

2. a certain truncated standard deviation of the $\hat{\theta}_a$ is set equal to one.

We will use the usual lower case symbols for parameters on this LOGIST scale. This should not cause confusion, since the lower-case parameters of Sections 1-3 will not be needed again.

If we start with LOGIST $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$, and $\hat{\theta}_a$ and determine k and K so that $\hat{\bar{B}}_0 = 0$ and $\hat{\bar{B}}_1 = 1$, then the $\hat{A}_i$, $\hat{B}_i$, $\hat{C}_i$ ($i = 1,2,\ldots,n$), and the $\hat{\Theta}_a$ ($a = 1,2,\ldots,N$), calculated by substituting estimated values into (15), will still satisfy the likelihood equations. The upper-case parameter estimates so obtained should have the sampling variance-covariance matrix found theoretically at the end of Section 4. Our remaining task is to compare an empirically determined variance-covariance matrix of MLE's with the corresponding theoretical matrix.

6. Recapitulation

We have used, at different points, three different arbitrary scales for our parameters:

1. θ_N and θ_{N-1} are assigned arbitrarily.
2. The origin is set at $\bar{B}_0$, the unit is $\bar{B}_1$.

3. The origin is set at a truncated mean of the θ_a , the unit is a truncated standard deviation of the θ_a .

Scale 1 (denoted by lower-case symbols) is most convenient mathematically for the difficult task of inverting the M -by- M information matrix. Scale 1 is not useful for practical purposes, however, since its use grossly inflates all the sampling variances.

Scale 2 (denoted by upper-case symbols) seems the simplest choice in an attempt to keep the sampling error in the estimated origin and unit as small as possible. The sampling variances computed for scale 1 are transformed (see eq. 19) to values appropriate for scale 2. Although scale 2 is not the familiar one, the two item sets used to specify the scale can be chosen so that the numerical values of $\hat{A}_i$, $\hat{B}_i$, $\hat{C}_i$ differ little from the familiar $\hat{a}_i$, $\hat{b}_i$, and $\hat{c}_i$ produced by LOGIST.

Scale 3 (hereafter denoted by lower-case symbols) is the scale used by LOGIST.

7. Empirical Estimation Procedures

As already stated, our theoretical results can be trusted only if they are shown to be in reasonable agreement with empirical results. For this purpose, artificial data $\|u_{ia}\|$ were created representing the administration of a 45-item test to a random sample of 1500

examinees. The 1500 θ_a were a spaced sample drawn from a distribution of abilities from a regular test administration. Six replicate matrices of $\|u_{ia}\|$ were independently generated, using the same item parameters and the same 1500 θ_a . The variation in responses across these matrices thus represents random fluctuations in u_{ia} for fixed a_i , b_i , c_i and θ_a .

Further replication was also built in: items 16-30 and items 31-45 had the same item parameters as items 1-15. The true lower-case and upper-case item parameters are shown in Table 1 for items 1-15.

Six independent runs were made on LOGIST, one for each group of 1500 examinees. For each run separately, $\hat{\bar{b}}_0$ was calculated from items 4-9, 19-24, 34-39; $\hat{\bar{b}}_1$ was calculated from items 10-15, 25-30, 40-45. It is convenient for our ultimate interpretation of the standard errors to be obtained that the true $\bar{b}_1 - \bar{b}_0 = .671 - (-.305) = .976$. Since this is close to 1.0, the scale unit for the capitalized parameters is very close to the scale unit for the lower-case (LOGIST) parameters.

For each run separately, all lower-case parameter estimates were linearly transformed as in (15) to the upper-case scale, using estimated k and K values. For the data reported in subsequent sections, the true $k = .976$ and the true $K = .312$. Since the six runs are independent, an unbiased empirical estimate of the sampling variance of any parameter estimate T is given by

Table 1

True (Upper Case) Item Parameters

Item No.	A	a	B	b	C or c
1	.96	.99	-1.75	-2.01	.17
2	.34	.35	-1.33	-1.61	.17
3	1.34	1.38	-.80	-1.09	.17
4	.76	.78	-.48	-.77	.17
5	.41	.42	-.38	-.67	.17
6	.90	.92	-.04	-.34	.17
7	.90	.92	.16	-.15	.17
8	1.04	1.06	.31	.00	.17
9	1.31	1.34	.42	.11	.13
10	1.46	1.50	.58	.26	.34
11	.85	.87	.79	.46	.17
12	.60	.62	.90	.57	.17
13	1.06	1.09	1.01	.68	.25
14	1.36	1.39	1.23	.90	.29
15	1.46	1.50	1.50	1.16	.18

$$s_{\hat{T}}^2 \equiv \frac{6}{5} \left[\frac{1}{6} \sum^{6} \hat{T}^2 - \left(\frac{1}{6} \sum^{6} \hat{T} \right)^2 \right] \tag{20}$$

the sum being across the six LOGIST runs. If the $\hat{T}$ in (20) were normally distributed, $s_{\hat{T}}^2/\sigma_{\hat{T}}^2$ would have an F distribution with 5 and ∞ degrees of freedom.

Since three different items have identical item parameters, the $s_{\hat{T}}^2$ for a single item parameter can be averaged across these three items to yield the best available unbiased estimate:

$$\bar{s}_{\hat{T}}^2 \equiv \frac{1}{3} \sum^{3} s_{\hat{T}}^2 \quad . \tag{21}$$

Note that it would be incorrect to pool all 18 values of $\hat{T}$ in an equation like (20), since $\hat{T}$ from the same LOGIST run are not independent.

If T_i and S_i represent two different item parameters in the same item

$$\bar{s}(\hat{T}_i, \hat{S}_i) \equiv \frac{1}{3} \sum^{3} s(\hat{T}_i, \hat{S}_i) \tag{22}$$

which is the same as (21) except that covariances are substituted for variances. If $\hat{T}_i$ and $\hat{S}_j$ represent item parameters in different items, then there are nine different sample covariances to be summed:

$$\bar{s}(T_i, \hat{S}_j) = \frac{1}{9} \sum^{3}\sum^{3} s(\hat{T}_i, \hat{S}_j) \quad . \tag{23}$$

If T is an ability parameter, (20) still holds. For our purposes, replacing T by θ, we can write

$$\bar{s}^2_{\hat{\theta}} = \frac{1}{N_g} \sum^{N_g} s^2_{\hat{\theta}} \tag{24}$$

where the sum is over all examinees in group g. When θ is at the midpoint of interval g, this average should be roughly equal to the $\sigma_{\hat{\theta}}$ obtained in Section 4.

If subscripts a and b denote different examinees in group g,

$$\bar{s}(\hat{\theta}_a, \hat{\theta}_b) = \frac{2}{N_g(N_g - 1)} \sum_{a>b}\sum s(\hat{\theta}_a, \hat{\theta}_b) \tag{25}$$

where the sum is over all pairs of examinees in group g. If a and b denote examinees in groups g and h respectively ($g \neq h$), then

$$\bar{s}(\hat{\theta}_a, \hat{\theta}_b) = \frac{1}{N_g N_h} \sum_{a=1}^{N_g} \sum_{b=1}^{N_h} s(\hat{\theta}_a, \hat{\theta}_b) \quad . \tag{26}$$

Finally, if T_i is an item parameter and examinee a is in group g, then

$$\bar{s}(T_i, \theta_a) = \frac{1}{3N_g} \sum^{3} \sum^{N_g} s(T_i, \theta_a) \quad . \tag{27}$$

In computing (24) - (27), examinees are grouped on their true values, not on their estimated values.

A problem arises when an examinee obtains a perfect score or a zero score. In this case his $\hat{\theta}$ is infinite and cannot be advantageously used. Instead of making some ad hoc adjustment, the 17 examinees for whom this occurred were simply removed from the group of examinees studied, leaving $N = 1483$. This has the effect of slightly biasing $\bar{s}$'s for the remaining most extreme θ values.

8. Numerical Standard Errors

Since the c parameter of an easy item usually cannot be accurately estimated, LOGIST in ordinary use does not estimate them individually. This would prevent the empirical standard errors of Section 7 from agreeing with the theoretical standard errors of Section 4. Since our main purpose is to show that the method of Section 4 can give useful results, the empirical and theoretical standard errors reported here are all estimated or calculated under the condition that the true values of c_i are known for $i = 1,2,3,4,5,12$. Items 1-5 are easy items, item 12 was included because of its low a_i . For empirical work, the true c values were supplied to LOGIST, which held them fixed while estimating all other parameters. For theoretical work, the rows and columns of (7) corresponding to c_1 , c_2 , c_3 , c_4 , c_5 ,

and c_{12} were simply deleted from the information matrix (7) before inversion.

Table 2 compares the empirical standard errors of Section 7 for $\hat{B}$ with the theoretical standard errors of Section 4. The last three columns show the squared ratios for the three replications of each item; each of these ratios will have an F distribution with 5 and ∞ degrees of freedom provided i) $\hat{B}$ has a normal sampling distribution, ii) $\hat{B}$ is unbiased, and iii) the theoretical $\sigma_{\hat{B}}$ from Section 4 is correct. An F above 2.21 or below .229 is significant at the (two-tailed) 10 percent level. Eleven of the ratios are significant. The number of ratios less than 1 is approximately the same as the number of ratios greater than 1.

In the past, the only available standard errors for item parameters assumed that the were known. Such standard errors for $\hat{B}$, for known θ, are given in the second column of the table. A comparison of second and third columns shows very close agreement except for the three easiest items (1,2,3). For these three items, our new theoretical value is larger and agrees better with the empirical value. This gives support to the new theoretical values. The fact that the empirical values (from Section 7) tend to be larger than the theoretical (from Section 4) could be due to n and N not being large enough for asymptotic results. A second likely explanation is that LOGIST was not really run to complete convergence.

Table 3 makes comparisons for $\hat{A}$. Again the standard errors of $\hat{A}$ with θ unknown agree closely with the results when is known. The empirical standard errors, although correlating well with the theoretical, seem to be larger. Eleven of the F ratios are

Table 2

Theoretical and Empirical Standard Errors for $\hat{B}$

Item No.	$\sigma_{\hat{B}\mid C}$ (C known)	$\sigma_{\hat{B}}$ (Sect. 4)	$\bar{s}_{\hat{B}}$ (Sect. 7)	$s^2_{\hat{B}}/\sigma^2_{\hat{B}}$		
1*	.110	.156	.183	.23	.56	3.34+
2*	.186	.201	.237	1.76	1.49	.93
3*	.045	.071	.063	1.38	.59	.41
4*	.060	.068	.066	.90	.76	1.17
5*	.100	.099	.103	.37	.40	2.48+
6	.125	.121	.131	.28	.63	2.63+
7	.113	.110	.100	1.24	.65	.58
8	.084	.083	.088	2.31+	.97	.16+
9	.055	.055	.067	.37	2.63+	1.47
10	.069	.069	.106	3.19+	3.62+	.33
11	.100	.097	.122	1.45	2.55+	.70
12*	.094	.091	.087	.85	1.27	.66
13	.086	.083	.094	1.01	1.20	1.57
14	.077	.076	.111	1.19	1.49	3.75+
15	.072	.075	.093	.40	2.62+	1.63

+Significant at 10 percent level.

*The C parameter for these items is treated as known.

Table 3

Theoretical and Empirical Standard Errors for $\hat{A}$

Item No.	$\sigma_{\hat{A}\vert\theta}$	$\sigma_{\hat{A}}$	$\bar{s}_{\hat{A}}$	$s^2_{\hat{A}}/\sigma^2_{\hat{A}}$		
1*	.088	.105	.141	.95	.91	3.60+
2*	.044	.046	.039	.88	.51	.74
3*	.097	.117	.094	1.39	.32	.22+
4*	.060	.065	.080	.89	2.77+	.86
5*	.045	.047	.054	.63	2.44+	.93
6	.103	.102	.123	1.54	.30	2.51+
7	.105	.105	.147	1.30	2.25+	2.35+
8	.113	.115	.159	1.29	3.20+	1.29
9	.123	.128	.182	1.89	3.39+	.80
10	.184	.193	.160	.71	.55	.79
11	.115	.120	.132	1.42	1.85	.34
12*	.060	.060	.076	.95	2.94+	.94
13	.151	.157	.187	2.40+	1.08	.79
14	.209	.218	.240	1.32	.91	1.43
15	.222	.233	.182	.25	.65	.93

+Significant at 10 percent level.

*The C parameter for these items is treated as known.

significant. Similar statements apply to Table 4, which shows the comparisons for $\hat{C}$.

Table 5 compares standard errors for $\hat{\theta}$. Let us leave column 3 for later discussion. Columns 4 and 5 show standard errors of $\hat{\theta}$ corresponding to the θ value in the first column; column 6, however, is computed from (2) for the group of N_g people falling in the class interval with midpoint $\hat{\theta}$. There is good agreement between empirical and theoretical standard errors except for $\theta < -1.5$. For low θ , asymptotic results do not appear with the usual n and N .

Table 5 shows close agreement of our standard error from Sections 2-4 with the standard error of $\hat{\theta}$ when the item parameters are known. The agreement shown here and in previous tables suggests that (1) is a good approximation to the diagonal of (2) and similarly for item parameters, that (2) agrees well with the empirical standard errors.

A comparison of the third and fifth columns in Table 5 shows what happens to σ when all C_i must be estimated from the data: For $\theta < -1$, σ is sharply affected; for $0 < \theta < 2.5$, there is very little effect.

Table 6 contains the squared ratios of the empirical standard errors to the theoretical standard errors for the five θ closest to the midpoint of the intervals, and within at least .1 of the midpoint. Two of the groups had only two abilities within this restriction. If similar caveats apply as for the item parameters these ratios will have an F distribution with five and ∞ degrees of freedom. Only eight of the ratios are significant at the two-tailed 10% level, and only 16 are greater than 1.

Table 4

Theoretical and Empirical Standard Errors in $\hat{C}$

Item No.*	$\sigma_{\hat{C}}(\infty)$	$\sigma_{\hat{C}}$	$\bar{s}_{\hat{C}}$	$s^2_{\hat{C}}/\sigma^2_{\hat{C}}$		
6	.056	.058	.063	.39	.44	2.79+
7	.049	.050	.038	.40	.35	.95
8	.037	.037	.045	3.08+	.76	.43
9	.024	.025	.039	.80	4.71+	1.83
10	.025	.026	.034	2.24+	2.68+	.27
11	.036	.037	.043	.98	2.67+	.41
13	.026	.027	.037	.89	1.88	2.90+
14	.019	.020	.028	2.98+	2.55+	.43
15	.015	.015	.016	.64	1.23	1.71

+Significant at 10 percent level.

*$C_1, \ldots, C_5$, and C_{12} are treated as known.

Table 5

Theoretical and Empirical Standard Errors for [illegible]

		All C_i unknown	C_1 to C_5 and C_{12} treated as known		
[illegible]	N_g	[illegible]	[illegible]$\|A,B,C$	[illegible]	[illegible]
-2.75	10	2.090	.951	.966	*
-2.25	35	1.296	.686	.699	1.134
-1.75	93	.861	.516	.525	.797
-1.25	219	.607	.400	.404	.427
-.75	332	.456	.341	.342	.332
-.25	326	.349	.295	.295	.279
.25	227	.278	.262	.263	.274
.75	136	.261	.260	.261	.286
1.25	77	.303	.289	.290	.349
1.75	25	.422	.384	.387	.412
2.25	3	.628	.575	.580	*
2.75	0	.931	.874	.878	*

*Not computed because of small N_g .

Table 6

F Ratios for $\hat{\theta}$

θ	$s^2_{\hat{\theta}}/\sigma^2_{\hat{\theta}}$				
-2.75	3.73†	4.41†			
-2.25	.85	.78	.43	11.34†	1.16
-1.75	.57	1.90	1.62	.32	18.95†
-1.25	.98	.63	.96	.95	.77
-.75	.26	.94	.63	.81	.63
-.25	.71	1.81	.73	.04†	.48
.25	.18†	.98	.74	.80	.77
.75	.61	.35	1.41	1.21	.64
1.25	2.76†	1.82	.98	1.08	1.84
1.75	.67	.41	1.08	1.45	1.78
2.25	.11†	.36			
2.75*					

†Significant at 10 percent level.

*There were no θ between 2.65 and 2.85.

Table 7 presents the theoretical standard errors of $\hat{A}$, $\hat{B}$, and $\hat{C}$, obtained by the method of Sections 2-4, when all C_i must be estimated from the data. It is interesting to compare these values with those in Tables 2-4 where $C_1,\ldots,C_5$, and C_{12} were treated as known. We find that the standard errors of $\hat{B}_1$ to $\hat{B}_5$ are increased drastically by ignorance of C_1 to C_5 ; all other $\sigma(\hat{B}_i)$ are much increased, except for $i = 11$, 13, and 14. All $\hat{A}_i$ show sharply increased standard errors. For items for which C_i must be estimated, on the other hand, the standard errors of $\hat{C}_i$ are little affected by knowledge or ignorance of $C_1,\ldots,C_5,C_{12}$. A likely explanation for this is that errors in estimating the scale unit B_1 affect the standard errors of the $\hat{A}_i$ and the $\hat{B}_i$, but not of the $\hat{C}_i$.

We have found in Tables 2-7 some illustrative answers to the question: How do estimation errors on one set of items affect the accuracy of estimated parameters for a different set of items? Such effects could not be quantified until now since the standard error of an item parameter estimate was previously known only for fixed θ . It is only through the sampling fluctuations of $\hat{\theta}$ that estimation errors for one item can affect parameter estimates for another item.

With 18 C_i treated as known, the Fisher information matrix inverted for this study has $3 \times 45 - 18 + 1498 = 1615$ rows and columns. The matrix inversion by the method of Section 4 used 1232K bytes of memory on an IBM 3031 and took 32 seconds. The computer program dealt with a 45-item test; it did not take advantage of the fact that the 45 items consisted of 3 replicate sets of 15 items each.

Table 7

Standard Errors (2) of Item Parameters when

All C_i Must be Estimated

Item No.	σ_B	σ_A	σ_C
1	.52	.23	.60
2	2.54	.13	.72
3	.35	.32	.10
4	.26	.15	.14
5	.97	.10	.32
6	.19	.18	.07
7	.16	.18	.06
8	.14	.21	.041
9	.12	.26	.026
10	.11	.32	.026
11	.10	.18	.039
12	.18	.14	.07
13	.09	.23	.027
14	.08	.31	.020
15	.10	.33	.015

In order to verify the numerical accuracy of the inversion, the information matrix and the variance-covariance matrix were multiplied. The result was an identity matrix accurate to 10 decimal places. The variance-covariance matrix obtained in double precision agreed with the matrix obtained in quadruple precision to all six decimal places printed.

9. Sampling Covariances and Correlations

When item parameters are known, $\hat{\theta}_a$ and $\hat{\theta}_b$ ($a \neq b$) are uncorrelated. When ability parameters are known, estimated item parameters for different items are uncorrelated. When both item and ability parameters are estimated, in general all estimates are correlated. The computer printout of the sampling correlations for the present study consists of 10 correlation matrices. These need only be summarized here.

Table 8 shows the theoretical (T) and empirical (E) correlations between estimates of two different parameters for the same item. The correlations are generally substantial. For comparison, the theoretical correlations when the abilities are known are included. The empirical correlations are obtained by dividing the estimated sampling covariance by the square roots of the estimated sampling variances. If the empirical correlations here have roughly 15 degrees of freedom, their standard error is roughly $(1 - \rho^2)/\sqrt{15} = .26(1 - \rho^2)$. In view of their standard errors, there is very satisfactory agreement of empirical with theoretical correlations.

Table 9 shows both theoretical and empirical correlations for the $\hat{B}_i$ ($i = 1,2,\ldots,15$). The corresponding standard errors are

Table 8

Theoretical (*T*) and *Empirical* (*E*) Sampling Correlations Between

Two Parameter Estimates for the Same Item

Item	$\rho_{\hat{A}\hat{B}\mid\theta}$	$\rho_{\hat{A}\hat{B}}$		$\rho_{\hat{B}\hat{C}\mid\theta}$	$\rho_{\hat{B}\hat{C}}$		$\rho_{\hat{A}\hat{C}\mid\theta}$	$\rho_{\hat{A}\hat{C}}$	
		T	E		T	E		T	E
1	.82	.86	.87						
2	.80	.82	.88						
3	.55	.70	.65						
4	.42	.52	.76						
5	.35	.38	.44						
6	.73	.70	.53	.92	.90	.92	.76	.70	.53
7	.67	.64	.66	.90	.88	.77	.76	.71	.79
8	.56	.52	.26	.83	.81	.85	.72	.67	.58
9	.37	.33	.50	.69	.67	.87	.65	.60	.81
10	.41	.42	.68	.69	.68	.93	.61	.61	.74
11	.40	.42	.70	.75	.74	.89	.77	.77	.83
12	-.55	-.51	-.79						
13	.22	.21	.06	.60	.59	.66	.69	.70	.67
14	.06	.03	.35	.45	.42	.61	.58	.59	.68
15	-.19	-.25	-.81	.25	.21	-.18	.53	.54	.56

TABLE 9
EXPERIMENTAL (E) AND THEORETICAL (T) STANDARD ERRORS (DIAGONALS) AND CORRELATIONS FOR TRANSFORMED B (DECIMAL POINTS OMITTED)

		B 1	B 2	B 3	B 4	B 5	B 6	B 7	B 8	B 9	B10	B11	B12	B13	B14	B15
B 1	E	(183)	141	284	045	193	-158	-028	-004	014	-122	289	016	-031	-064	-147
	T	(156)	264	509	334	158	-124	-128	-088	-040	034	-001	039	-005	-022	-050
B 2	E	141	(237)	541	286	-092	-036	126	-005	-252	-105	029	-064	-247	360	-040
	T	264	(201)	284	184	078	-066	-069	-046	-017	026	007	022	-002	-018	-039
B 3	E	284	541	(063)	308	040	-091	056	004	-274	-279	268	007	-298	348	-155
	T	509	284	(071)	377	151	-131	-139	-093	-032	048	008	044	-008	-032	-068
B 4	E	045	286	308	(066)	120	-072	-179	038	-362	-308	-007	192	-443	343	218
	T	334	184	377	(068)	066	-130	-130	-089	-040	029	003	028	-005	-018	-039
B 5	E	193	-092	040	120	(103)	-228	-126	-072	046	-205	236	086	-051	-193	126
	T	158	078	151	066	(099)	-117	-113	-088	-062	-004	-009	013	-003	001	002
B 6	E	-158	-036	-091	-072	-228	(131)	014	076	-041	107	-153	002	016	122	-085
	T	-124	-066	-131	-130	-117	(121)	-062	-053	-051	-016	004	-005	001	004	011
B 7	E	-028	126	056	-179	-126	014	(100)	-120	098	121	-050	-221	156	-018	000
	T	-128	-069	-139	-130	-113	-062	(110)	-042	-036	-011	002	-009	002	005	011
B 8	E	-004	-005	004	038	-072	076	-120	(088)	-068	025	-015	101	-062	081	-137
	T	-088	-046	-093	-089	-088	-053	-042	(083)	-007	004	001	-010	002	002	003
B 9	E	014	-252	-274	-362	046	-041	098	-068	(067)	198	037	-129	332	-357	-063
	T	-040	-017	-032	-040	-062	-051	-036	-007	(055)	023	-003	-013	002	000	-005
B10	E	-122	-105	-279	-308	-205	107	121	025	198	(106)	-193	-137	270	-151	-098
	T	034	026	048	029	-004	-016	-011	004	023	(069)	-035	-052	-043	-062	-087
B11	E	289	029	268	-007	236	-153	-050	-015	037	-193	(122)	041	-011	-103	-182
	T	-001	007	008	003	-009	004	002	001	-003	-035	(097)	-071	-067	-086	-107
B12	E	016	-064	007	192	086	002	-221	101	-129	-137	041	(087)	-176	078	005
	T	039	022	044	028	013	-005	-009	-010	-013	-052	-071	(091)	-069	-068	-065
B13	E	-031	-247	-298	-443	-051	016	156	-062	332	270	-011	-176	(094)	-341	-112
	T	-005	-002	-008	-005	-003	001	002	002	002	-043	-067	-069	(083)	-057	-060
B14	E	-064	360	348	343	-193	122	-018	081	-357	-151	-103	078	-341	(111)	006
	T	-022	-018	-032	-018	001	004	005	002	000	-062	-086	-068	-057	(076)	-004
B15	E	-147	-040	-155	218	126	-085	000	-137	-063	-098	-182	005	-112	006	(093)
	T	-050	-039	-068	-039	002	011	011	003	-005	-087	-107	-065	-060	-004	(075)

given in parentheses in the diagonal. The only theoretical correlations above .20 are among $\hat{B}_1$, $\hat{B}_2$, $\hat{B}_3$, and $\hat{B}_4$. These are the four easiest items. Any error in estimating the scale unit $\bar{B}_1 - \bar{B}_0$ would seriously affect all these items in the same way. It is hard to draw other useful generalizations from this table.

The corresponding table for the $\hat{A}_i$ ($i = 1,2,\ldots,15$) shows only 3 theoretical correlations above .20: $\rho_{13} = .27$, $\rho_{14} = .20$, $\rho_{34} = .23$. With two exceptions ($\rho_{67} = -.013$, $\rho_{6,12} = -.002$), all theoretical correlations are positive.

The highest theoretical correlation among the $\hat{C}_i$ ($i = 6,7,\ldots,$ 11 and 13, 14, 15) is $\rho_{67} = .04$. All correlations are positive.

The theoretical correlations between $\hat{A}_i$ and $\hat{B}_j$ ($i \neq j$) are all below .20 in absolute value, except for items 1-4, which vary from .14 to .38. For $\hat{B}_i$ and $\hat{C}_j$ ($i \neq j$; $j \neq 1,2,\ldots,5,12$) there are no correlations above .25 in absolute value. For $\hat{A}_i$ and $\hat{C}_j$, there are no correlations above .20 in absolute value.

The theoretical correlations between $\hat{}_a$ and $\hat{}_b$ ($a \neq b$) are all less than .04 in absolute value. Between $\hat{}_a$ and $\hat{B}_i$, the largest correlation in absolute value is .15 (when $i = 1$ and $= -2.25$). Between $\hat{}_a$ and $\hat{A}_i$, the largest is .12 (when $i = 1$ and $= -2.25$). Between $\hat{}_a$ and $\hat{C}_i$, the largest is .06.

Summary

When both abilities and item parameters are unknown, the asymptotic sampling variance-covariance matrix developed in this paper appears to provide useful values for the standard errors needed for further research in item response theory. The magnitude of the numerical values in the matrix were very much affected by the method used to define the scale. For a set of artificial data, this variance-covariance matrix compared satisfactorially with empirical results; also with the variance-covariance matrices found by the usual formulas for the case where the abilities are known or where the item parameters are known.

With this matrix, the effect on other items of including items with poorly determined parameters can be studied. Including items with poorly determined c 's increases the standard errors of all of the a 's and b 's but not of the other c 's. The effect of different distributions of abilities on the accuracy of item parameters can also be studied. Hopefully a goodness-of-fit test can now be developed for the three-parameter model.

The standard errors of item parameters can now be studied for a situation of common occurrence in equating and item banking: Each of two tests containing common items is administered to a different group of examinees; all parameters are estimated in the same LOGIST run. It is of particular interest to determine how the number of common items affects the standard error of the parameter estimates.

References

Bradley, R. A. & Gart, J. J. The asymptotic properties of ML estimators when sampling from associated populations. Biometrika, 1962, 49, 205-214.

Haberman, S. J. Maximum likelihood estimates in exponential response models. The Annals of Statistics, 1977, 5, 815-841.

Kendall, M. G. & Stuart, A. The advanced theory of statistics (Vol. 1, 3rd ed.). New York: Hafner, 1969.

Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

DISTRIBUTION LIST

Navy

1 Dr. Jack R. Borsting
Provost and Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Chief of Naval Education and
Training Liason Office
Air Force Human Resource Laboratory
Flying Training Division
Williams Air Force Base, AZ 85224

1 CDR Mike Curran
Office of Naval Research
800 North Quincy Street
Code 270
Arlington, VA 22217

1 Dr. Pat Federico
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. Paul Foley
Navy Personnel R & D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R & D Center
San Diego, CA 92152

1 Patrick R. Harrison
Psychology Course Director
Leadership and Law Department (7b)
Division of Professional Development
U.S. Naval Academy
Annapolis, MD 21402

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CNV-70)
Newport News Shipbuilding and
Drydock Co.
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. William Nordbrock
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

1 Library, Code P201L
Navy Personnel R & D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Building 114, Section D
666 Summer Street
Boston, MA 02210

1 Office of Naval Research
Code 437
800 North Quincy Street
Arlington, VA 22217

5 Personnel and Training Research Programs
Code 458
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research Development and Studies Branch
OP-115
Washington, DC 20350

1 The Principal Deputy Assistant Secretary of the Navy (MRA&L)
4E780, The Pentagon
Washington, DC 22203

1 Director, Research and Analysis Division
Plans and Policy Department
Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203

1 Mr. Arnold Rubenstein
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 Dr. Worth Scanland, Director
Research, Development, Test and Evaluation
N-5
Naval Education and Training Command
NAS
Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Department of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. J. B. Sympson
Naval Personnel R & D Center
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Services
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R & D Center
San Diego, CA 92152

1 Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. John H. Wolfe
Code P310
U.S. Navy Personnel Research
and Development Center
San Diego, CA 92152

1 Mr. Ted M. I. Yellen
Technical Information Office
Code 201
Navy Personnel R & D Center
San Diego, CA 92152

Army

1 Technical Director
U.S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 LTC Michael Plummer
Chief, Leadership and Organizational
Effectiveness Division
Office of the Deputy Chief of Staff
for Personnel
Department of the Army
The Pentagon
Washington, DC 20301

1 Dr. James L. Raney
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for
the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U.S. Army Research Institute for
the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commandant
U.S. Army Institute of Administration
Attn: Dr. Sherrill
Ft. Benjamin Harrison, IN 46256

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Laboratory
AFHRL/MPD
Brooks Air Force Base, TX 78235

1 U.S. Air Force Office of
Scientific Research
Life Sciences Directorate
Bolling Air Force Base
Washington, DC 20332

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks Air Force Base, TX 78235

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling Air Force Base
Washington, DC 20332

1 Dr. David R. Hunter
AFHRL/MOAM
Brooks Air Force Base, TX 78235

1 Research and Measurement Division
Research Branch, AFMPC/MPCYPR
Randolph Air Force Base, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks Air Force Base, TX 78235

Marines

1 Dr. H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower
Utilization
HQ, Marine Corps (MPU)
BCB, Building 2009
Quantico, VA 22134

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

1 MAJ Michael L. Patrow, USMC
Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

1 Dr. A. L. Slafkosky
Scientific Advisor
Code RD-1
HQ, U.S. Marine Corps
Washington, DC 20380

Coast Guard

1 Chief, Psychological Research Branch
U.S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Mr. Thomas A. Warm
U.S. Coast Guard Institute
P.O. Substation 18
Oklahoma City, OK 73169

Other DoD

1 DARPA
1400 Wilson Boulevard
Arlington, VA 22209

12 Defense Technical Information Center
Cameron Station, Building 5
Attn: TC
Alexandria, VA 22314

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Director, Research and Data
OASD (MRA&L)
3B919, The Pentagon
Washington, DC 20301

1 Military Assistant for Training
and Personnel Technology
Office of the Under Secretary of
Defense for Research and Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. Wayne Sellman
Office of the Assistant Secretary
of Defense (MRA&L)
2B269 The Pentagon
Washington, DC 20301

Civil Government

1 Mr. Richard McKillip
Personnel R & D Center
Office of Personnel Management
1900 E Street, NW
Washington, DC 20415

1 Dr. Andrew R. Molnar
Science Education Development
and Research
National Science Foundation
Washington, DC 20550

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and
Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Vern W. Urry
Personnel R & D Center
Office of Personnel Management
1900 E Street, NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory and Cognitive Processes
National Science Foundation
Washington, DC 20550

Non-Government

1 Dr. James Algina
University of Florida
Gainesville, FL 32611

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 Psychological Research Unit
Department of Defense (Army Office)
Campbell Park Offices
Canberra, ACT 2600
AUSTRALIA

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08541

1 CAPT J. Jean Belanger
Training Development Division
Canadian Forces Training System
CFTSHQ, CFB Trenton
Astra, Ontario KOK 1B0
CANADA

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
ISRAEL

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 3
D-5300 Bonn 2
WEST GERMANY

1 Dr. R. Darrell Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Liaison Scientists
Office of Naval Research
Branch Office, London
Box 39
FPO, NY 09510

1 Dr. Robert Brennan
American College Testing Programs
P.O. Box 168
Iowa City, IA 52240

1 Dr. C. Victor Bunderson
WICAT Inc.
University Plaza, Suite 10
1160 S. State Street
Orem, UT 84057

1 Dr. John B. Carroll
Psychometric Laboratory
University of North Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts and Sciences
University of Rochester
River Compus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Department of Psychology
University of Southern California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N
Washington, DC 20036

1 Dr. Fritz Drasgow
Yale School of Organization and
Management
Yale University
Box 1A
New Haven, CT 06520

1 Dr. Mike Durmeyer
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. A. J. Eschenbrenner
Dept. E422, Bldg. 81
McDonald Douglas Astronautics Co.
P.O. Box 516
St. Louis, MO 63166

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray and Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Psychologisches Institut der
Universitat Wien
Liebiggasse 5/3
A 1010 Wien
AUSTRIA

1 Prof. Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organization
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt, Beranek, and Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Robert Glaser
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Daniel Gopher
Industrial and Management Engineering
Technion-Israel Institute of
Technology
Haifa
ISRAEL

1 Dr. Bert Green
Department of Psychology
Johns Hopkins University
Charles and 34th Streets
Baltimore, MD 21218

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Delwyn Harnisch
University of Illinois
242b Education
Urbana, IL 61801

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Jack Hunter
2122 Coolidge Street
Lansing, MI 48906

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Prof. John A. Keats
Department of Psychology
University of Newcastle
Newcastle, New South Wales 2308
AUSTRALIA

1 Mr. Jeff Kelety
Department of Instructional Technology
University of Southern California
Los Angeles, CA 90007

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Building
University of Illinois
Champaign, IL 61801

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
NETHERLANDS

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands, Western Australia 6009
AUSTRALIA

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08541

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Prof. Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurement
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army Navy Drive
Arlington, VA 22202

1 Dr. Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Diane M. Ramsey-Klee
R-K Research and System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Mr. Minrat M. L. Rauch
P II 4
Bundesministerium der Verteidigung
Postfach 1328
D-53 Bonn 1
GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Department
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 Dr. Andrew Rose
American Institutes for Research
1055 Thomas Jefferson St., NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 Prof. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Robert Sternberg
Department of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. Patrick Suppes
Institute for Mathematical Studies in
the Social Sciences
Stanford University
Stanford, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massacuusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 East River Road
Minneapolis, MN 55455

1 Dr. Susan E. Whitely
Psychology Department
University of Kansas
Lawrence, KS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

END

DATE
FILMED

11-82

DTIC